FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate

L. M. Moreno
M. A. Mansilla
S. A. Bullard

Please see article for additional authors.
Nonsyndromic orofacial clefts are a common complex birth defect caused by genetic and environmental factors and/or their interactions. A previous genome-wide linkage scan discovered a novel locus for cleft lip with or without cleft palate (CL/P) at 9q22–q33. To identify the etiologic gene, we undertook an iterative and complementary fine mapping strategy using family-based CL/P samples from Colombia, USA and the Philippines. Candidate genes within 9q22–q33 were sequenced, revealing 32 new variants. Concurrently, 397 SNPs spanning the 9q22–q33 2-LOD-unit interval were tested for association. Significant SNP and haplotype association signals \( P = 1.45 \times 10^{-8} \) narrowed the interval to a 200 kb region containing: FOXE1, C9ORF156 and HEMGN. Association results were replicated in CL/P families of European descent and when all populations

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
were combined the two most associated SNPs, rs3758249 \((P = 5.01E \times 13)\) and rs4460498 \((P = 6.51E \times 12)\), were located inside a 70 kb high linkage disequilibrium block containing FOXE1. Association signals for Caucasians and Asians clustered 5’ and 3’ of FOXE1, respectively. Isolated cleft palate (CP) was also associated, indicating that FOXE1 plays a role in two phenotypes thought to be genetically distinct. Foxe1 expression was found in the epithelium undergoing fusion between the medial nasal and maxillary processes. Mutation screens of FOXE1 identified two family-specific missense mutations at highly conserved amino acids. These data indicate that FOXE1 is a major gene for CL/P and provides new insights for improved counseling and genetic interaction studies.

**INTRODUCTION**

Isolated cleft lip with or without cleft palate (CL/P) is a common complex birth defect that results from genetic variations, environmental exposures and/or their interactions. Family-based studies have estimated that CL/P genetic risk involves 3–14 loci interacting multiplicatively (1). Isolated cleft palate (CP) has often been separated in analysis from CL/P based on embryologic and recurrence-risk evidence although overlap in causal genes has been observed in some syndromic forms of clefting (2). Multicenter collaborative studies, utilizing both linkage and association approaches applied to substantial numbers of multiplex and nuclear triad families, have significantly advanced knowledge of locus, gene and environmental exposures to CL/P etiology (3–12). Discoveries include the strong effect of common IRF6 variants (9), the collective occurrence of rare DNA variations in 20 genes: TGFβ3, MSX1, PTCH, FOXE1, GLI2, JAG2, LHX8, MSX1, MSX2, SATB2, SKI, SPRY2, TBX10, PVRL1, FGFR1, FGFR2, FGFR8, CRISPLD2, BMP4 and RYK (13–23) and significant association or linkage to various genes (24). Most recently published genome-wide association scans for NSCL/P have identified a major locus at 8q24.21 (25,26) that confers a substantial risk in European populations.

Maternal periconceptional smoking is a well-established environmental risk factor (27); alcohol use and poor nutrition have also been suggested as risks factors for oral clefts but the data for these are weaker (28). In addition, gene-environment interactions in 20 genes (6,29–31).

To identify additional CL/P risk genes, we conducted a genome-wide linkage scan on 388 multiplex families from seven populations, revealing highly significant evidence of linkage to the 9q22–q33 region with an HLOD of 5.5 (11) (Fig. 1). Potential candidate genes based on gene expression or known role in human clefting syndromes mapping within the region include ROR2 (104 cM), FGD3, BARX1, PTCH, FOXE1, GABAB2R, TGFβ1 and ZNF189 (120 cM).

We describe here the detailed fine mapping of the 9q22–q33 region, using parallel positional cloning and candidate gene strategies, leading to the discovery that common FOXE1 locus variants have a significant role in the etiology of nonsyndromic orofacial clefting.

**RESULTS**

**Stage 1: 2-LOD-unit support interval association mapping**

Analysis of the Illumina SNP panel revealed the strongest association results within FOXE1 from Colombia and USA_1 families (for rs1443434, \(P = 8.30E \times 04\) and \(P = 6.69E \times 03\), respectively) (Fig. 2A and Supplementary Material, Appendix Table S1). Other association signals were noted within the IARS gene in Colombians and PALM2 in Filipinos. Upon combining all three populations (combined data), rs1443434 located in the FOXE1 3’-UTR returned the most significant evidence of association \((P = 3.8E \times 05)\), indicating the presence of a common variant within the approximately 70 kb long haplotype block containing FOXE1.

**Stage 2: candidate genes**

Convergent association to FOXE1 was also observed through analysis of the candidate genes within the initial linkage peak, (Fig. 2B, and Supplementary Material, Appendix Table S2). Two SNPs, rs894673 and rs3758249, within 3.7 kb 5’ of FOXE1 contributed the strongest results in both the Colombian and USA_1 populations \((P\)-values range from 8.67E \times 03 to 1.8E \times 04). Although combining the Filipino data set with these two populations increased the significance to \(P\)-values <E \times 06 at both rs894673 \((P = 8.25E \times 07)\) and rs3758249 \((P = 8.64E \times 07)\) the strongest signal for the Filipinos was within the PTCH gene at rs2297088 \((97.28\ Mbp, P\text{-value} = 6.49E \times 03)\), (Fig. 2B, and Supplementary Material, Appendix Table S2). Of note, over-transmission of the PTCH rs2297088–rs2236407 haplotype with borderline statistical significance \((P = 0.08)\) in Filipino families with two or more affected members has been reported previously (16).

**Stage 3: the critical region XPA-HEMGN**

Given the strong association to a region at most 357 kb in size flanked by the XPA and HEMGN genes and containing FOXE1, additional SNPs were genotyped to further narrow the critical region. Reconstruction of linkage disequilibrium (LD) blocks in all populations (Fig. 3A–C) showed similar LD patterns for Colombia and USA, while LD blocks of smaller size were observed for the Filipinos (Fig. 3C). Association results excluded the XPA gene for all populations, refining the critical region to 200 kb \((rs1877431–rs4618817; 99.57–99.77\ Mbp)\) containing FOXE1, C9ORF156 and
Figure 1. Maximum summed multipoint HLODs on chromosome 9 under a dominant model obtained from the 10 cM genome scan of CL/P in the CIDR-7 studies. Each population’s contribution to the total HLOD is detailed elsewhere (11,68).

Figure 2. FBAT association results. (A) 331 Illumina markers including the 2LOD interval. (B) Candidate genes interval between ROR2-ALDOB (178 SNPs). (C) 357 kb critical region (34 SNPs). The size of the circles represents the over-transmission of the common (large circles) or rare allele (small circles).
Figure 3. LD Blocks for Stage 3—357 kb critical region SNPs. FBAT association results $-\log_{10}(P\text{-value})$ are displayed by upper track. (A) Colombia, (B) the USA, (C) the Philippines.
HEMGN (Fig. 2C). The highest significance occurred within a 70 kb block (rs10217225–rs6586; 99.634–99.707 Mbp) encompassing FOXE1 for the Colombian (rs4460498, \( P = 2.3E - 05 \)) and USA_1 families (rs894673, \( P = 1.12E - 03 \)) (Fig. 2C and Supplementary Material, Appendix Table S3). Less significant results were observed in the Filipinos (rs4460498, \( P = 1.58E - 03 \)) possibly a reflection of low minor allele frequencies. Combining all three populations produced highly significant results with markers rs894673 (\( P = 7.27E - 08 \)), rs3758249 (\( P = 7.64E - 08 \)), rs1867280 (\( P = 7.49E - 08 \)), and rs4460498 (\( P = 3.73E - 08 \)).

**Stage 3: haplotype analysis**

For the Colombian and the USA_1 samples, the most significant haplotypes were clustered in a region spanning 9.5 kb immediately 5’ of FOXE1, with some haplotypes extending into the gene (Fig. 4A and B). In addition, moderately significant haplotypes were found between rs1877431 and rs16924274 at −82 to −59 kb upstream and another between rs993501 and rs6586 spanning 43.5 kb starting 7.2 kb downstream of FOXE1. The most significant Filipino haplotypes occurred between rs874004 and rs872251, a 135.2 kb region starting 6 kb 3’ of FOXE1 (Fig. 4C). Given several associated haplotypes overlapped between all three data sets, populations were combined yielding increased significance beyond each population alone, confirming that haplotypes immediately 5’ of FOXE1 showed the strongest association (\( P = 1.45E - 08 \)) (Fig. 4D).

**Stage 4: replication panel analysis**

Markers surrounding FOXE1 (\( n = 15 \)), included in haplotypes with significant results for the Colombian and USA_1 populations, were genotyped in an independent set of CL/P samples that included subjects from Denmark, Norway and Iowa. Results obtained for Denmark (rs3021523 \( P = 2.18E - 03 \)) and Norway (rs4460498 \( P = 4.96E - 03 \)) showed independent replication of the initial FOXE1 association signal (Supplementary Material, Appendix Table S6). Interestingly, rs4460498 was also the most strongly associated marker for the Colombian and Filipino data sets (\( P = 2.3E - 05 \) and 1.90E – 03, respectively). Iowa samples were pooled with the additional USA_1 samples, and the results (USA_2) were significant for rs3758249 (\( P = 5.60E - 05 \); Bonferroni correction = 1.67E – 04), replicating the association with markers in FOXE1 (Supplementary Material, Appendix Table S6).
Figure 4. Haplotype window analysis results for Stage 3—357 kb critical region. Rectangles represent the size of the haplotype according to SNPs involved. Gray scale represent the $-\log_{10}(P$-value) results. (A) Colombia (B) USA (C) Philippines (D) Combined. See Supplementary Material, Appendix Table S1 for allele correspondence to specific nucleotides.
Table 1. Number of pedigrees by cleft phenotypic subgroup

<table>
<thead>
<tr>
<th>Population</th>
<th>Clefting subgroups</th>
<th>CLO</th>
<th>CLPO</th>
<th>CL/P</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colombia</td>
<td>28</td>
<td>214</td>
<td>272</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>102</td>
<td>175</td>
<td>324</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>69</td>
<td>178</td>
<td>298</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>191</td>
<td>200</td>
<td>395</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>122</td>
<td>192</td>
<td>314</td>
<td>118</td>
<td></td>
</tr>
</tbody>
</table>

Italicized subgroups are overlapping. Clefting subgroups: CLO, pedigrees with cleft lip only affected individuals; CLPO, pedigrees with CLP only affected individuals; CL/P, pedigrees in which members are affected with at least a cleft lip; CP, pedigrees with at least one cleft palate only affected individual.

**FOXE1 analysis by phenotype**

To test the hypothesis that FOXE1 has a role in isolated CP, DNA from cases with CP (Table 1) were genotyped for the 15 SNPs surrounding FOXE1. Association with CP was observed in the USA, Filipino, Norwegian and Danish families (Caucasian data set did not have CP families) and when combining all populations, rs1867278 was the most significant (\( P = 4.1 \times 10^{-4} \)) (Fig. 5E), indicating that FOXE1 is also involved in the etiology of isolated CP.

To explore whether the association for the CL/P sample was due to a specific subset type of cleft, the entire data set was analyzed across all cleft phenotypes (Fig. 5A–E). The CLO subset consists of those families in which one or more of the affected family members have cleft lip alone, while all affected family members in the CLPO subset have cleft lip and cleft palate. The CL/P subgroup is a combination of the CL and CLPO subgroups, plus families with both CL and CLP. The families in the CP subgroup have at least one affected individual with isolated cleft palate. CLPO was more significantly associated in the USA and Norwegian families, while the Colombian and Danish families showed similar association for both CLO and CLPO. Of note, the Colombian data set had 10-fold fewer CLO families than CLPO families. In contrast, no association was found with Filipino CLO families (Fig. 5B and C, Supplementary Material, Appendix Table S6). When all populations were combined the same markers, rs3756249, rs1867278 and rs10217225, combined the most significant results occurred at rs3758249 (\( P = 5.01 \times 10^{-13} \)), rs1867278 (\( P = 7.44 \times 10^{-13} \)) and rs4460498 (\( P = 6.15 \times 10^{-12} \)). The over-transmission percentage was highest to T-G-T-G-A-C-C, all common alleles except rs10217225, that was the most significant in Colombian, USA, Norwegian and Danish families. Another one located 21 kb 3′ and spanning 120 kb consisting of rs10119795–rs10984103–rs7853349–rs6586–rs4618817–rs872251 (2-2-2-1-1-1, corresponding to T-C-T-T-A-A) was the most significant in Filipinos (Figs 4 and 7). This latter haplotype was also observed in the other populations albeit with less significance.

**Allele dosage effects**

A trend towards a dosage effect was seen for the presence of two copies relative to one copy of the over-transmitted alleles for each marker (rs3758249 and rs4460498), for both associated phenotypes (CL/P and CLPO; see Supplementary Material, Appendix Table S7). However, as the overlapping confidence intervals indicate (Supplementary Material, Appendix Table S7), this difference is not statistically significant.

**Parent of origin effects**

Marked parent of origin effects were seen for the Colombian and Philippines population with both CL/P (Fig. 8 and Supplementary Material, Appendix Table S4) and CLPO (not shown) phenotypes and these FOXE1 alleles. Over-transmission was seen preferentially from mothers compared with fathers. On the other hand, preferential over-transmission was observed among the fathers within the Caucasian data (USA and Scandinavian data sets) possibly indicating different roles of these alleles in different populations (Fig. 8).

**Sequencing results**

Sequencing of conserved non-coding regions in five candidate genes in probands from families that yielded positive LOD scores in the 9q22–33 region (11) (63 Filipino, 28 Colombians and 1 USA) identified 24 new variants. Sequencing of the coding regions in FOXE1 found eight new variants, two are missense mutations (I59S and P208R), six are synonymous changes (Supplementary Material, Appendix Table S8). Both missense mutations are predicted to be benign by PolyPhen (http://genetics.bwh.harvard.edu/pph/). All 32 variants are shown in Supplementary Material, Appendix Table S8 together with their allele frequencies, conservation and regulatory potential according to UCSC 5 way regulatory potential track.

Upon performing TESS searches, we found that 29 of these variants either disrupt or create a transcription factor binding site (TFBS). These analyses predicted 63 disrupted TFBS and 61 created TFBS (Supplementary Material, Appendix Table S9). From those we investigate further the top 25% according to a log likelihood/length of TFBS of 2 and log likelihood greater than 12 (Supplementary Material, Appendix Table S10). Four out of the 31 top TFBS signals are strong candidates for involvement in craniofacial development:
Figure 5. FBAT results. (A) All cleft phenotypes. (B) CLO by population. (C) CLPO by population. (D) CL/P by population and (E) CP by population.
Expression results

Foxel expression was observed at E10.5 and E11.5 in the rostral epithelium of the oral pharynx, including the caudal epithelium of the medial nasal and maxillary processes (Fig. 9). In particular, Foxel is expressed in the epithelium involved in the fusion between the medial nasal and maxillary processes. Neither C9orf156 nor Hemgn were observed during primary palatogenesis (data not shown). These data are consistent with Foxel playing a role in lip development.

DISCUSSION

Through multiple stages involving positional cloning, candidate gene sequencing and developmental gene expression analysis, we have identified FOXE1 as a major disease gene within the previously published 9q22–33 linkage peak. An initial panel of 300 SNPs narrowed the 20 Mb 2-LOD interval to 357 Kb. Fine mapping data that excludes the two genes (XPA and HEMGN) flanking this interval and expression data, negative for C9orf156 and positive for Foxel, converge on FOXE1 as the etiologic gene.

Within a 200 kb critical region, a 20 kb disease haplotype with greater effect in Colombian and Scandinavian samples was identified that extends 19.2 kb upstream of the FOXE1
start codon (Fig. 7). Although all populations analyzed in this study contributed to the FOXE1 association signal, SNP and haplotype frequencies found within each population indicate that this predisposing variant is more common in populations of Caucasian descent. This is in contrast to the IRF6 predisposing SNP rs642961 G/A (9,32) that is more common in populations of Asian descent. Our data indicate two additional FOXE1 risk alleles. Independent signals from a haplotype spanning 95.5 kb starting 21.1 kb 3' of FOXE1 in our Filipino population indicate an Asian-specific allele (Figs 4C and 7). Previous reports of a translocation breakpoint 55 kb 3' of FOXE1 in two siblings with bilateral cleft lip and palate lend credence to this hypothesis (33). Filipinos have demonstrated similar evidence for two or more variants contributing to clefting at the IRF6 locus (9,32). Finally, the presence of a third allele in a separate LD block is supported by significant SNPs and haplotypes containing rs1443432 and rs780258 67 kb upstream of FOXE1 (Figs 3–6). Finding multiple risk alleles are to be expected for a complex disease such as orofacial clefting.

The association patterns suggest the phenotypic spectrum of FOXE1 variants involves all combinations of primary and secondary palatal clefts. These findings differ from those of the IRF6 rs642961 variant disrupting an AP-2α-binding site that was not associated with CP in Filipinos. Furthermore, in contrast to IRF6, FOXE1 markers could not distinguish CLPO or

Figure 6. Over-transmission of the associated allele. (A) CL/P and (B) CLPO.
From CLO in our populations. Both isolated cleft lip and isolated cleft palate can be separated from, or lumped together with patients having CLPO based on embryology, recurrence and syndromic gene findings. FOXE1 appears to present one unifying path for these phenotypes.

Of note is the maternal over-transmission of FOXE1 alleles in the Colombian and Filipino samples compared with all other Caucasian samples as well as the paternal over-transmission observed in families from the USA and Scandinavia. This may reflect a parent of origin effect such as imprinting, although there is no existing evidence to suggest imprinting for the FOXE1 locus. Preferential maternal over-transmission could also indicate a maternal genetic effect. There is dissimilarity between the maternal and paternal ancestral lineages of Colombian families with a predominance of Native American mitochondrial haplotypes (87.5%) and Caucasian Y chromosome haplotypes (57.4%), indicating admixture events between Caucasian immigrant men and Native American women (34); a common finding amongst South American populations (35,36). A Brazilian orofacial cleft study not only demonstrated similar differences between ancestral maternal and paternal contributions, but also showed significantly higher frequency of Native American mitochondrial and Y haplotypes in cases compared with controls (37). Furthermore, evidence of genetic heterogeneity due to Native American ancestry was underscored by the presence of association between CL/P and IRF6 ($P = 0.023$) and of CL with RFC1 ($P = 0.017$) in non-carriers of the Native American haplogroup D (38).

Similar parent of origin effects for orofacial clefting have been described for CBS, IRF6, MTHFR, MTR, PAX3, PAX7, RUNX2, TGFA and TGFB3 (6,39–45). Most interesting were the results reported by Shi et al. (6), indicating a maternal genetic effect for IRF6 with the fetal risk allele being protective when the mother is a risk allele carrier.

Only two FOXE1 coding mutations, I59S and P208R, were identified in our screen of over 200 patients. Neither mutation is present in 24 control individuals from this study or in 186 controls in a candidate gene mutation screen (15). P208R is immediately adjacent to the A207V identified in a patient with CL/P (15). I59S is located in the first alpha helix of the forkhead DNA-binding domain within which two mutations, S57N and A65V, have been described in families with features of the Bamforth–Lazarus syndrome (BLS) (46,47). The S57N and A65V mutations have been shown to significantly reduce DNA binding and transcription activation. Furthermore, an I84S mutation at the corresponding FOXL2 amino acid as FOXE1 I59 has been reported in a family with blepharophimosis-ptosis-epicanthus inversus syndrome (48). Therefore, the I59S mutation in our study may have similar effects on FOXE1 function. The lack of a more severe phenotype in our proband may be due to only one mutated FOXE1

Figure 7. Haplotype window analysis results for replication panel markers and the CL/P phenotype. Rectangles represent the size of the haplotype according to SNPs involved. Colors represent the $-\log_{10}(P$-value) results. i, rs#, markers typed by CIDR; rs#, markers genotyped in our laboratories. See Supplementary Material, Appendix Table S11 for allele correspondence to specific nucleotides.
copy, whereas BLS is recessive. The phenotypic spectrum of FOXE1 coding mutations may include features less severe than BLS, such as hypothyroidism. Of note, FOXE1 variants confer a 5.7-fold for thyroid cancer (49).

The variants identified by sequencing the noncoding regions of FOXE1, PTCH, TGFBR1, GABABR2, ZNF189, FGD3 could presumably have etiologic effects. Although none were associated with CL/P, they occurred at conserved nucleotides and some were predicted to affect TFBSs, including SP1, AP1, AP3 and POU1F1 that are involved in facial development.

FOXE1 is a member of a transcription factor family that contains a DNA-binding forkhead domain and that are involved in embryonic pattern formation. Newborn mice null for Foxel exhibit cleft palate and thyroid anomalies (50). 

Fcoll is expressed in the secondary palate epithelium in both mice at E13.5–E15.5 (51) and humans at week 11 of gestation (52). The specific expression pattern of Foxel at the point of fusion between maxillary and nasal processes, described for the first time in this study, strongly suggests FOXE1 as a key player in primary palatogenesis. 

Foxel is a downstream target of the Shh/Gli pathway in hair follicle morphogenesis (53). Thus, it is plausible that there are two disease genes in the q22–q33 region since the SHH receptor PTCH was associated with CL/P in our Filipino data set and there is ample evidence that SHH signaling is involved during primary palatogenesis (54). Furthermore, there is overlapping Foxel and Shh expression in the caudal epithelium of the mesial nasal and maxillary processes. 

The paucity of FOXE1 mutations despite exhaustive searches (15) and in light of the highly significant association described here indicate that causal mutations are in nearby noncoding regions that regulate FOXE1 expression, most likely within the 70 kb high LD block containing the 5′-UTR and promoter. Recently, a variant affecting MYF-5 DNA binding in a cluster of Gli-binding sites 1.2 kb upstream of the transcription start site has been described in 11 of 25 patients with NSCLP (55). We found this variant in 9 of 184 case and 2 of 186 control chromosomes. While this is coincident with etiologic effects, this variant is too infrequent to explain the strong association results unless it splits a common haplotype into two disease alleles.

Efforts to find other causal variants will require in depth sequencing and characterization of regulatory function in highly conserved regions within our putative disease haplotypes. Utilizing the Vertebrate Multiz Alignment (28 species) and the PhastCons track of the UCSC browser (56), we performed a preliminary screening of the 200 kb critical region with a set threshold score above 350. Fourteen regions with high PhastCons scores (Table 2) were identified. One region that contains the most significantly associated SNP also harbors five Gli2-binding sites (57). The first of these Gli2-binding sites is located 20 bp downstream of rs3758249, suggesting further a connection of FOXE1 to the Shh/Gli pathway for primary palatogenesis.

In the aggregate, our findings place FOXE1 as the third locus along with IRF6 and 8q24 in which common variants have a substantial impact on the occurrence of cleft lip and palate. It affords a new pathway to investigate biologically and new tools to improve genetic counseling.

MATERIALS AND METHODS

Subjects

Inclusion criteria for participating subjects were a diagnosis of nonsyndromic clefting (58). Subjects were clinically examined, medical and family histories reviewed and gestational environmental exposures were analyzed to rule out syndromes and known teratogens. The Colombian families comprise a collection of extended families and nuclear triads from the province of Antioquia in Northwest Colombia and were recruited through the Clinica Noel in Medellin (273 affected pedigrees plus 90 control triads). The USA Caucasian families were recruited as part of a Multicenter Collaborative Project and include families from the states of California, Washington, Ohio and Iowa (464 pedigrees). Filipino pedigrees (577 total) were ascertained through the Operation Smile Organization (59). Independent replication panels comprised of families from Denmark, Norway and Iowa. Samples from Denmark were 592 case triads consisting of affected individuals (born 1981–1990), their parents and some unaffected sibling(s). A Danish cohort of unaffected control individuals was also included, yielding 499 genotyped children (born 1997–2003), 444 typed mothers, and 313 adults. Samples from Norway (60) were triads recruited as part of a case-cohort study of facial clefts (born 1996–2001), including 452 case triads and 762 control triads. Finally, the samples from Iowa consisted of a collection of 249 case triads and 207 unrelated controls.
For analysis purposes, the Iowa case sample was pooled with all other USA samples (Table 3).

Cleft phenotypic subgroups were determined from the cleft history of first-, second- and third-degree relatives of the proband. Analysis of markers in FOXE1 was stratified by cleft phenotype in subsets of pedigrees of the following cleft phenotypes: cleft lip only families (CLO) where all affected family members present a cleft lip only. Families with cleft lip and palate only (CLPO) are those in which all affected family members have a CLP phenotype. The CL/P phenotypic group includes all pedigrees in which individuals had either a CLO or CLPO. Finally, CP pedigrees are those where there is at least one individual affected with a cleft palate only among other possible cleft phenotypes. An overall group of families (ALL group) represents the combination of CL/P and CP groups, and also includes those few pedigrees with a nonsyndromic cleft of unknown type (Table 1).

All samples utilized in our study were obtained after individuals and/or their parents gave signed informed consent under a protocol approval by institutional review boards at the Scientific Committee at the Dental School of University of Antioquia-Colombia, the Ohio State University, the University of Iowa, Children’s Hospital and Regional Medical Center in Seattle, the Hope Foundation (Bacolod City, Negros-Philippines), the University of Pittsburgh, and the ethics committees in Denmark and Norway.

Genotyping and association analysis

Parallel fine mapping and candidate gene approaches were employed iteratively using association and sequencing methods to identify the locus responsible for the 9q22–q33 linkage signal.

In the first approach (Stage 1), the region between 102 and 140 cM including the 2-LOD-unit support interval was evaluated with 331 SNPs genotyped in 571 pedigrees with CL/P from Colombia, USA_1 and the Philippines (Table 3). The genotyping was performed by the Center for Inherited Disease Research (CIDR) using a custom Illumina SNP panel. The goal of the initial fine mapping was to saturate all known genes in the 1-LOD-unit support interval (Fig. 1) with 2–6 SNPs per gene while filling intergenic gaps of >100 kb with 1 SNP/20 kb. The -1 to 2-LOD-unit support interval was covered in a similar fashion; however, intergenic regions had approximately 1 SNP/50 kb. To cover the second Colombian peak (120–140 cM) (9), clusters of 2–4 SNPs every 200–400 kb were seeded across this interval. This saturation provided adequate SNP resolution in the Paisa population from Colombia (62). SNPs were chosen using the information from the Hapmap Project (63), http://www.hapmap.org, based on the CEPH (Utah residents with ancestry from northern and western Europe) (CEU), Japanese in Tokyo, Japan (JPT) and Han Chinese in Beijing, China (CHB) populations. Tagging SNPs were evaluated in regards to minor allele frequency, deviations from Hardy–Weinberg equilibrium (HWE), inter-marker distance, as well as LD patterns and haplotype block structures using HAPLOVIEW version 2.05 (56). Tagging SNPs were selected to include at least 85% of the haplotypes in the LD blocks and MAF was set to 20% minimum (for complete list of SNPs, see Supplementary Material, Appendix Table S1).

In parallel, candidate genes within the 2-LOD-unit support interval (Stage 2) (ROR2, FDG3, BARX1, PTCH, FOXE1, GPR51, COL15A1, TGFBR1, ZNF189 and ALDOB) were tested for association by genotyping 44 additional SNPs genotyped with TaqMan assays from Applied Biosystems (Foster City, CA, USA) in the same panel of both Colombian and USA pedigrees used in Stage 1 plus 26 additional pedigrees from USA and 307 Filipino triads. Genotypes obtained from 134 Illumina SNPs (from Stage 1) located within these selected candidate genes were included in the association analysis (for complete list of SNPs, see Supplementary Material, Appendix Table S2).

In Stage 3, a critical region of 357 kb (99.44–99.8 Mbp) containing the genes XPA, FOXE1, C9ORF156 and HEMGN was fine mapped by another round of genotyping and association analysis using a total of 34 SNPs (22 new Taqman assays, 5 Taqman assays genotyped in stage 2 and 7 Illumina SNPs from stage 1) (Supplementary Material, Appendix Table S3).

Single SNP association was tested using family-based association methods (FBAT) (64) utilizing the –e option to account for multiple sibs in a family or multiple nuclear families in a pedigree. HBAT haplotype analysis was used for genotypic data on Stage 3 (critical region, XPA-FOX1E1-C9ORF156-HEMGN) using a sliding window approach with a maximum of five SNPs per window.
Table 2. Regions harboring potentially regulatory elements

<table>
<thead>
<tr>
<th>Priority</th>
<th>Name relative to ATG</th>
<th>Position start (UCSC March 2006 Assembly)</th>
<th>Position end (UCSC March 2006 Assembly)</th>
<th>Size (bp)</th>
<th>PhastCons placental mammal scores</th>
<th>Distance relative to ATG (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CRS +20.7</td>
<td>99 676 550</td>
<td>99 677 150</td>
<td>600</td>
<td>615</td>
<td>20 550</td>
</tr>
<tr>
<td>2</td>
<td>CRS +67</td>
<td>99 588 450</td>
<td>99 588 900</td>
<td>450</td>
<td>604</td>
<td>-67 550</td>
</tr>
<tr>
<td>3</td>
<td>CRS +9.4</td>
<td>99 665 300</td>
<td>99 665 750</td>
<td>450</td>
<td>514</td>
<td>9300</td>
</tr>
<tr>
<td>4</td>
<td>CRS -50</td>
<td>99 605 600</td>
<td>99 605 950</td>
<td>350</td>
<td>507</td>
<td>-50 400</td>
</tr>
<tr>
<td>5</td>
<td>CRS -51.5</td>
<td>99 604 125</td>
<td>99 604 900</td>
<td>775</td>
<td>440</td>
<td>-51 875</td>
</tr>
<tr>
<td>6</td>
<td>CRS +23.0</td>
<td>99 678 800</td>
<td>99 679 700</td>
<td>900</td>
<td>398 390 396</td>
<td>22 800</td>
</tr>
<tr>
<td>7</td>
<td>CRS +4</td>
<td>99 660 000</td>
<td>99 660 400</td>
<td>400</td>
<td>525</td>
<td>4000</td>
</tr>
<tr>
<td>8</td>
<td>CRS -82</td>
<td>99 573 500</td>
<td>99 574 700</td>
<td>900</td>
<td>425</td>
<td>-82 200</td>
</tr>
<tr>
<td>9</td>
<td>CRS -44.2</td>
<td>99 611 750</td>
<td>99 612 050</td>
<td>300</td>
<td>492</td>
<td>-44 250</td>
</tr>
<tr>
<td>10</td>
<td>CRS -22.4</td>
<td>99 633 350</td>
<td>99 634 050</td>
<td>700</td>
<td>474</td>
<td>-22 650</td>
</tr>
<tr>
<td>11</td>
<td>CRS -9.0</td>
<td>99 647 200</td>
<td>99 647 450</td>
<td>250</td>
<td>460</td>
<td>-8800</td>
</tr>
<tr>
<td>12</td>
<td>CRS -72.0</td>
<td>99 583 600</td>
<td>99 584 000</td>
<td>400</td>
<td>379 357 360</td>
<td>-72 400</td>
</tr>
<tr>
<td>13</td>
<td>CRS -2.2</td>
<td>99 653 800</td>
<td>99 654 450</td>
<td>650</td>
<td>4 GLI binding sites</td>
<td>-2200</td>
</tr>
<tr>
<td>14</td>
<td>CRS -0.45</td>
<td>99 654 950</td>
<td>99 655 350</td>
<td>400</td>
<td>419</td>
<td>-1050</td>
</tr>
</tbody>
</table>

Selected regions for searches of potential regulatory elements based on PhastCons placental mammal scores above 350, 7X regulatory potential, transcription binding sites. Region 13 shows no PhastCons placental mammal scores but contains four GLI binding sites and association results. CRS, conserved regulatory sequence.

Table 3. Number of pedigrees and markers included in the study

<table>
<thead>
<tr>
<th>Stage</th>
<th>331 SNPs</th>
<th>178 SNPs</th>
<th>34 SNPs</th>
<th>15 SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 LOD INTERVAL</td>
<td>ROR2-ALDOB</td>
<td>XPA-HEMN</td>
<td>FOXE1</td>
</tr>
<tr>
<td>Case Populations</td>
<td>331 Illumina</td>
<td>134 Illumina</td>
<td>44 Taqman</td>
<td>7 Illumina</td>
</tr>
<tr>
<td>Colombia</td>
<td>232</td>
<td>232</td>
<td>232</td>
<td>232</td>
</tr>
<tr>
<td>USA_1 California, Washington and Ohio</td>
<td>130</td>
<td>130</td>
<td>156</td>
<td>130</td>
</tr>
<tr>
<td>USA_2 California, Washington, Ohio and Iowa</td>
<td>209</td>
<td>209</td>
<td>209</td>
<td>209</td>
</tr>
<tr>
<td>Philippines multiplex families_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philippines multiplex families and trios_2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>592</td>
<td>592</td>
<td>592</td>
<td>592</td>
</tr>
<tr>
<td>Norway</td>
<td>452</td>
<td>452</td>
<td>452</td>
<td>452</td>
</tr>
<tr>
<td>TOTAL</td>
<td>571</td>
<td>362</td>
<td>695</td>
<td>362</td>
</tr>
</tbody>
</table>

USA_1 and USA_2 are collectively referring to in the text as the USA Caucasian families. (A) Cases (B) Controls.
In Stage 4, association with alleles of FOXE1 was replicated by further genotyping a cluster of 15 SNPs surrounding FOXE1 in an independent CL/P sample panel consisting of cohorts from Denmark \(n = 395\) peds with CL/P and Norway \(n = 314\) peds with CL/P as well as 139 additional CL/P affected triads and 207 unrelated controls from USA (Tables 1 and 3). To further evaluate the role of FOXE1 in different cleft phenotypes, the same 15 SNPs mentioned above were genotyped in additional CP pedigrees (Table 1, Supplementary Material, Appendix Table S6). All FOXE1 genotypic data were tested for association by performing SNP and haplotype TDT analyses as implemented in the FBAT software and subsetting by cleft phenotype (Table 1, Supplementary Material, Appendix Table S6) (64).

Correction for multiple testing was done by adjusting the alpha level of significance by means of Bonferroni corrections. \(P\)-values obtained in Stage 1 (2-LOD-unit support interval) analysis were not adjusted in an effort not to miss variants of lower effect. However \(P\)-values obtained in Stages 2 (178 SNPs) and 3 (34 SNPs) were adjusted using Bonferroni corrections with alpha levels of significance calculated as \(0.05/178 = 0.000281\) and \(0.05/34 = 0.00147\), respectively. In addition, \(P\)-values obtained for FOXE1 analyses by cleft phenotype were adjusted using a Bonferroni correction with an alpha level of significance calculated as \(0.05/300\) (15 SNPs \(\times\) 5 populations \(\times\) 4 cleft phenotypes, CLO, CLPO, CL/P, CP) = \(1.67E-04\).

Parental over-transmission

To investigate parent of origin effects within FOXE1, we performed overall parental TDT and parent-specific TDT calculations as implemented in S.A.G.E., which provides the exact McNemar test statistics for the familial data within the CLPO and CL/P phenotypic groups. Furthermore, the overall parental TDT transmission counts were utilized to calculate an over-transmission percentage rate. Assuming a null hypothesis of 50% transmission rate for any allele from heterozygous parents, the over-transmission rate was the ratio of the count of over-transmitted allele to the total count of transmitted alleles minus 50% (0.50 in decimal format). Ninety-five percent confidence intervals were calculated (assuming the over-transmission rate follows a binomial distribution) to determine if these over-transmission rates significantly differed among populations or cleft sub-groups. The over-transmission rates represent an alternative approach to the PAR% calculations to assess the overall impact of the associated allele.

Associated allele dosage effects

To test for dosage effects of the associated allele, the Weinberg LRT method was used to calculate the relative risk (RR) for heterozygotes and homozygotes for the over-transmitted allele for the most significantly associated SNP within cleft groups driving the association (65). Log-linear models with genotype and imprinting effects were fit to the data using the SAS version of the LRT program (65). Ninety-five percent confidence intervals were calculated to determine if the RR for carrying one copy of the associated allele differed significantly from the RR of two copies of the associated allele. Using the parameter estimate (\(b\)) and \(b\)'s 95% CI as the arguments to the exponential function, the RR and its 95% CI are calculated: \(RR (95\% CI) = e^{b\text{estimate}} (e^{\text{lower estimate}}, e^{\text{upper estimate}})\).

Sequencing

Previously, re-sequencing of coding regions in 9q22–33 candidate genes, PTCH, FOXE1 and TGFBR1, revealed only six private missense mutations (3, 2 and 1, respectively) in 180 CL/P patients (15,16). Therefore, in parallel with the association analysis above and in an effort to identify potential regulatory variants that could explain the linkage signal, 41 conserved non-coding regions between these and other clefting candidate genes (GABABR2, ZNF189, FGD3) were re-sequenced in 63 Filipino, 28 Colombian and 1 USA affected probands from families yielding positive 9q22–33 LOD scores. Selection criteria for non-coding regions to sequence were: at least 70% conservation in nucleotide identity over a minimum of 100 bp in five species (dog, mouse, rat, chicken and frog). A total of 34 kb of sequence was covered with 63 primer pairs (Supplementary Material, Appendix Table S5). Variants found were evaluated using the UCSC browser for 5X regulatory potential and conservation (http://genome.ucsc.edu/). In addition, to investigate the potential of creating or deleting TFBS, both the wild-type and variant allele were queried in the Transcription Element Search Software (TESS) (66) (http://www.cbil.upenn.edu/cgi-bin/tess/tess).

Upon narrowing the disease locus, FOXE1 5'-UTR, coding and 3'-UTR regions were also sequenced in the affected probands from families yielding positive 9q22–33 LOD scores described above. All sequencing reactions were performed as described previously (16).

Expression analysis

Expression of Fox1, Hemgn and C9orf156 during primary palatogenesis was characterized by in situ hybridization. C57BL/6J mouse embryos at E9.5, E10.5 and E11.5 were evaluated by both whole mount and section in situ with gene-specific digoxigenin-labeled RNA probes (67). Fox1 (NM183298) probe is a 0.8 kb BamHI/NorI cDNA clone in pBluescript (generously provided by Roberto DiLauro). Hemgn (NM.053149) probe is a 529 bp cDNA clone in pCRII-TOPO. C9orf156 homolog clone 5830415F09 (BC092543) is a 560 bp clone in pCRII-TOPO. Coordinates according to the UCSC Mm July 2007 and primer sequences for both Hemgn and C9orf156 probes are as follows: mmHemgn-ex34F, AGA-TGT-CGC-TGA-AGG-CTG-TC 46409176–46409195; mmHemgn-ex34R, TTC-CGG-ATC-TTG-ATC-GTC-CT 46407541–46407560; mmC9orf156-ex34F, CAA-GCT-GGA-GAA-GTG-AG 46398966–46398985; mmC9orf156-ex34R, GGA-AGC-ATC-CCA-CTG-TGT-TC 46395064–46395083.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.
ACKNOWLEDGEMENTS

We are especially grateful to the families that participated in this study and also the staff at each recruiting site. Special thanks to the Universidad de Antioquia Paternity testing lab for providing access to anonymous Colombian control samples. We would like to thank Xiaoyan Qin, Peng Shi, Sandy Imoehl, Scott Pate and Jason Dierdorff for their help with genotyping experiments and Kristin Oleson for sample preparation. Thanks to Susie Mc Connell, Nancy Davin and Daniel Benton for their administrative support.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Institutes of Health [R01-DE014667-8, KO2-AEO15291, a March of Dimes Basil O’Connor award # FY 98-0718 and Research Grant #6-FY01-616, a Cleft Palate Foundation Grant and an American Association of Orthodontics Foundation Faculty Development Award to A.C.L., DE-08559 to J.C.M., DE-11948 to K.C., PS0-DE-16215 to M.L.M. and J.C.M., R01-DE016148 to M.L.M., R21-DE016930 to M.L.M., R01-DE09886 to M.L.M., R01-DE012472 to M.L.M. and Fogarty 1RO3-TW-007644 to J.C.M.]; and also by the Intramural Research Program of the National Institute of Environmental Health Sciences, National Institutes of Health. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, Contract Number N01-HG-65403. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Dental and Craniofacial Research, or the National Institutes of Health.

REFERENCES


